Proposed Data Set for Bidding for the Part D Drug Benefit

Issue

After preliminary discussions with the American Academy of Actuaries (AAA), CMS is proposing to make data available to facilitate the preparation of bids by potential drug plan sponsors for the provision of Part D prescription drug coverage. In order to further inform our plans for this data set, CMS would like to discuss our tentative plan with a larger audience to further refine the specifications of data to be provided. This paper provides background information on data that CMS currently plans to make available.

Background

Medicare currently does not include an outpatient prescription drug benefit. As a result, we do not have Medicare claims payment information for prescription drugs as we do for services under Parts A and B. We do, however, have statistically representative "micro-data" on the drug use and drug spending of Medicare beneficiaries, along with information about health, income and insured status through the Medicare Current Beneficiary Survey (MCBS).

CMS representatives discussed a number of options with representatives from the AAA. Some of the issues addressed included the following:

- Importance of CMS data to the bidding process
- What are the data needs for this purpose?
- Data elements of interest
- Years of data
- Availability of data (timing)
- Format and requirements for accessing the data

These elements will be discussed in the Open Door Forum. However, some of the more salient points are shown in Attachment I.

Options for Bidders Data Sets

Various options for data sets were considered by CMS. While there was an initial focus on the need to have multiple years of complete data discussants were concerned that many of the proposed data sets are extremely complex. That is, many users would not be able to manipulate these data sets. For example, many data systems would not be able to house the claims for 5% of the Medicare population (which are publicly available). In addition, it was also determined that some of the data sets such as the MCBS are technically very challenging to manipulate. (Some data are already available. They are described in Attachment 2 along with their web addresses.)

Another area of discussion focused on the need for certain elements of information that are not found in total in any of the available data sources. The process of imputation may assist in using multiple data sets to provide a more comprehensive data set for bidders.

As a result of the preliminary discussions, CMS is considering data releases that could include the following:

- Individual level data from the MCBS
- Continuance tables based on MCBS data
- Medicare 5% claims data with imputed drug use
- State level expenditure adjusters based on Federal retirees in a national plan

A description of each of these data sets is shown below.

1. Data from the MCBS

Data sets from the MCBS for public use are issued on a calendar year basis. The **Cost and Use** file for a given year is made available about 2 years after the close of fieldwork. This file contains reconciled information on events, charges, and payments from both survey and claims for the year. It also includes summaries of use and expenditures for the year from Medicare files along with survey data on insurance coverage, health status and functioning, and income. The sample represents the "ever-enrolled" Medicare population--all persons who were in the program at any time during the calendar year. (A description of the MCBS is shown in Attachment 3).

We anticipate releasing a subset of the elements available in the 2001 MCBS Cost and Use file – the most recent such file available. We expect that these elements will include demographic characteristics such as age and sex, insurance characteristics (including existence of prescription drug coverage), summary Medicare Part A and Part B use and expenditure, and annual prescription drug expenditure (with a breakdown of this spending by various payer categories, including consumer payments). Note that the full dataset is available to researchers who enter into a data use agreement with the Agency.

This data file will be available in a Microsoft Excel spreadsheet, in a SAS (version 8) transport file, and in a pipe-delimited text file. The file will contain one record for each survey participant. Because the full 2001 Cost and Use file is already complete, the timing of the release of this extract file is constrained only by the decisions on which elements to include.

*Constraints on the data***:** We have identified three major weaknesses of this data for use by bidders on the Part D program**.** First, the data are not representative below the national level. The MCBS sample is drawn from the national Medicare enrollment list, so that survey participants are representative of that population. There are a number of states for which there are no geographic sampling areas. Even for those states with geographic sampling areas, the people in those areas are not necessarily representative of all enrollees in that State. For example, there may be only a single sampling area in a small

State, covering a rural district; participants probably are not representative of the urban enrollees in that State.

Second, the data are from 2001. Although this is the most recent data available of its sort (2002 data should become available late in calendar year 2004 or early in 2005), the data fail to capture recent pharmaceutical launches and recent movements of drugs to off-patent status.

Third, as with any person-based survey the MCBS results are subject to reporting error. Analysis of the 1999 Cost and Use file using a pharmacy-based survey suggested that drug expenditures were underestimated in the MCBS by an average of 17 percent, owing to participant recall errors. The evidence suggests that this proportion is higher among people with cheap or infrequent drug usage and lower among high-expenditure users or those using drugs to treat or control chronic conditions.

2. Distributions of Total Claims Costs (Data for Continuance Tables)

We attempted to address some of these weaknesses when we used MCBS data as the base from which to project calendar year 2006 prescription drug costs under Medicare Part D. The MCBS drug expenses were adjusted to correct for survey underreporting. The MCBS spending was brought forward to 2006. Since the MCBS cannot determine drug expenses for institutionalized beneficiaries, drug expenses for this group were imputed. Finally, the data were standardized to a full-retail cost level by removing the estimated effects of rebates and discounts.

We are proposing to provide distributions of prescription drug costs for Medicare beneficiaries based on our latest estimates (2004 Trustees Report). These distributions can be directly converted to continuance tables for analytical purposes. We expect to provide these distributions by December 2004.

We will provide the following distributions of drug expenses in selected dollar intervals for fee-for-service Medicare beneficiaries at projected 2006 expense levels.

- All beneficiaries (community and institutionalized)
- Community beneficiaries (non-institutionalized)

We plan to provide the fraction of people (percent of beneficiaries) and their mean expense in each dollar interval.

Although these distributions will provide information not otherwise available about Medicare beneficiaries' prescription drug expenses, users should not rely exclusively on these distributions for pricing determinations for a variety of reasons including…

- The MCBS is a survey with respondent recall error. The distributions will include our adjustments to compensate for these errors.
- The MCBS does not collect drug expenses for institutionalized beneficiaries. The distribution for all beneficiaries will include our imputed daily-institutionalized drug expenses.

- Per capita drug expenses are projected from the survey year to 2006 using our published National Health Expenditures projections.
- The MCBS drug prices vary by payer. The distributions will include our adjustments to the payments to reflect the effect of current discounts and rebates, and the expected drug plan discounts, rebates, and management.

3. Medicare 5% Claims Data with Imputed Drug Use

Description of Standard Analytic File (SAF): The SAF is available by type of claim or collectively as a group. The 5% sample is created based on selecting records with 05, 20, 45, 70 or 95 in positions 8 and 9 of the Health Insurance Claim (HIC) number. The DESY Link Key field provides an encrypted number that enables users to find all claims for a single beneficiary. These files contain final action claims data in which all adjustments have been resolved. This file is representative at the state level.

Imputation of drug use: Prescription drug data from the MCBS will be imputed to each beneficiary in the 5% sample file. This will be done using a statistical matching procedure. Records on both files will be assigned to cells based on characteristics that are present in both files and which prove to be good predictors of drug use. Possible matching criteria include age, gender, use of medical services, diagnoses, risk score, etc.. The drug information from one of the MCBS observations in a cell will be imputed to a beneficiary in the same cell on the 5% sample file. A geographic indicator, such as state code, will also be included. This procedure preserves the distribution of drug spending present in the MCBS whereas using a regression procedure would assign every observation in a cell the mean spending of the same cell on the MCBS.

The augmented file could be tabulated with more refined breaks of the estimates, such as state-level estimates. However, any sub-national estimates should be used with caution, as there is nothing in the process that will capture differences in practice patterns on the state level. Also, any time two different data sources are merged, a higher degree of error is introduced than is present in either sample by itself.

4. Geographic Prescription Drug Utilization Index

A set of relative indexes for per capita drug utilization measures has been prepared using measures from a data set of federal retires 65 and older in a single benefit plan with nationwide coverage. The benefit is uncapped; it has 2 levels of copayment for mail order and a coinsurance for store purchases. The geographic level of detail is the state.

There are indexes for per capita
1) Expenditures
2) Prescriptions filled
3) Days supplied

In addition, there are indexes for the above variables broken down by mail order and store purchases. Each index is normalized to 1.0 separately. There are no adjustments for demographics or morbidity in these measures.

Issues to be Addressed During the Open Door Forum

1. Will this data set be helpful to potential bidders?
2. What questions and concerns do potential bidders have?
3. What additional information do they need?
4. Is the proposed timing of the release of the data sets appropriate?
5. What other assistance can CMS provide to assist potential bidders?
6. Next steps

Attachment I
Issues Discussions

This list of items associated with bidding data was used to guide our discussions and discussion points made are described in more detail below.

Importance of the data to the bidding process

- There was a common understanding that data is critical to the bidding process. Although some organizations may have data that would be sufficient to support the bidding process, others will need additional data in order to develop responsible bids. Some organizations would find it very difficult to bid without any additional data.
- Information on all types of Medicare beneficiaries would be needed.
- Data on drug utilization and dollars in combination with population characteristics, such as, health insurance status, could be used to predict drug utilization for potential enrollee populations.
- Data provided at a more granular level, for example disaggregated down to the script level with therapeutic classes, could be used to analyze benefit structures. This would include simulating the results of changing copayments and deductibles.

What are the data needs for this purpose

- Data presented in both an aggregated level, such as continuance tables, or at a more micro level could be used for the purpose of bidding. Different organizations appear to have varying levels of need with regard to having access to aggregate and/or micro level data. Obviously, micro level data requires more resources in order to be analyzed.
  - There was great interest in continuance tables and preferably with very fine breaks in order to provide more information.
  - The extent of analysis is limited by the inability to dissect tables into finer cells. This would make it difficult to conduct analyses on various benefit structures. Bidders would therefore also like to have access to micro level data at the beneficiary and/or script level. This allows for analyses of prescription utilization by different population characteristics. It would also allow for simulations such as the substitution of generics for brands.

Data elements of interest

- Income status, age and sex
- Insurance status i.e. Medicare, medigap, employer coverage, Medicaid, uninsured, duals, and employer duals
- Geography and particularly by regions or states
- Expenditure information, but not necessarily pricing information
- Disabled status
- Script level information particularly counts, with brand vs. generic, mail vs. retail, NDC-9s, dosage and days supply

Years of Data

- Longitudinal information covering 3 years would be ideal in order to be able to conduct trend analyses, at a minimum for the population over age 65.
- Claims level information for one year would be sufficient.

Availability of data (timing)

- A preference for at least some form of preliminary data being made available in January, with final data sets being put out in March.
- Would need at least a 3-6 month lead time to work with the data – for June bidding

Format of and requirements for the data

- Excel formats are generally accessible to all but some data (i.e., the 5% sample) will need to be made available via flat files or other formats
- Data User Agreements are generally required for individual level data
- Fees may be applied to obtain these data

Attachment 2
Currently Available Data

In addition, currently available information may also provide helpful information. This data includes:

- HCC and other related data
  - Web address - http://www.cms.hhs.gov/healthplans/riskadj/
  - Data includes tentative diagnoses for the CMS-Hierarchical Condition Category risk adjustment model, average Part A/B factors by state – stlvlrskscores.zip (3K), national and county level counts and proportions of people with various risk factors - NatlFreq.zip (107K) and ffsdist1997edited.zip (3,616K)

Information on MCBS is provided at the following website:

- Web address - http://www.cms.hhs.gov/mcbs/DFDcnu00.asp/

Attachment 3
What is the MCBS?

The Medicare Current Beneficiary Survey (MCBS) is a continuous, multipurpose survey of a nationally representative sample of aged, disabled, and institutionalized Medicare beneficiaries. MCBS, which is sponsored by the Centers for Medicare & Medicaid Services (CMS), is the only comprehensive source of information on the health status, health care use and expenditures, health insurance coverage, and socioeconomic and demographic characteristics of the entire spectrum of Medicare beneficiaries.

Interviews at 4-month intervals are designed to yield longitudinal series of data. These data cover use of health services, medical care expenditures, health insurance coverage, sources of payment (public and private, including out-of-pocket payments), health status and functioning, and a variety of demographic and behavioral information such as income, assets, living arrangements, family supports, and access to medical care. An effort is made to interview the sampled person directly, but if the person is unable to answer the survey questions, he or she is asked to designate a proxy respondent, usually a family member or close acquaintance who is familiar with his or her care. On the average, 15 percent of the community interviews are done with proxy respondents.

The typical MCBS interview starts with a review of household composition and the sample person's health insurance coverage (including public program participation, Medicare supplementary policies, and HMO membership). The review also produces information on insurance premiums and sources of payment for those premiums. It is followed by a complete enumeration of health care utilization in the period "since the last interview;" survey participants are prompted to recall incidents of each type of service (hospital admission, physician visit, drug prescription/refill, etc.). Details of services provided, provider characteristics, and medicines prescribed are gathered, as is a detailed account of charges and payments associated with these health care events. These questions are asked each interview, so that the resulting database contains a continuous record of utilization, charges, and payments over multiple years for each sample person.

MCBS interview data are linked to Medicare claims and other administrative data to enhance their analytic power. This results in a database combining data that can be obtained only from personal interviews with Medicare administrative data. The survey and claims data together constitute a more complete utilization data set for the MCBS sample than is available from either source. Administrative data, such as buy-in status and capitated plan membership, are also added to the file. The final file consists of survey, administrative, and claims data. All personal identifying information is removed.